

MAP: Microblogging Assisted Profiling of TV Shows

Xiahong Lin¹, Zhi Wang², Lifeng Sun¹

¹ Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University

² Graduate School at Shenzhen, Tsinghua University
{lin-xh12@mails., wangzhi@sz., sunlf@}tsinghua.edu.cn

Abstract. Online microblogging services that have been increasingly used by people to share and exchange information, have emerged as a promising way to profiling multimedia contents, in a sense to provide users a socialized abstraction and understanding of these contents. In this paper, we propose a microblogging profiling framework, to provide a social demonstration of TV shows. Challenges for this study lie in two folds: First, TV shows are generally *offline*, i.e., most of them are not originally from the Internet, and we need to create a *connection* between these TV shows with online microblogging services; Second, contents in a microblogging service are extremely noisy for video profiling, and we need to strategically retrieve the most related information for the TV show profiling. To address these challenges, we propose a *MAP*, a microblogging-assisted profiling framework, with contributions as follows: i) We propose a joint user and content retrieval scheme, which uses information about both actors and topics of a TV show to retrieve related microblogs; ii) We propose a social-aware profiling strategy, which profiles a video according to not only its content, but also the social relationship of its microblogging users and its propagation in the social network; iii) We present some interesting analysis, based on our framework to profile real-world TV shows.

Keywords: Online Microblogging, TV Show Profiling, Social Network, Data Visualization

1 Introduction

Recent years have witnessed an increasing popularity of online microblogging services (e.g., Twitter and Weibo), and the rapid convergence between the online microblogging service and other multimedia services including online video streaming [13]. This new trend makes the online microblogging service a promising way to *profiling* multimedia contents, i.e., using the information from the microblogging systems to abstract, demonstrate and enrich multimedia contents [6]. In this paper, we study profiling one of the most important multimedia types — TV shows (videos) that are generally received by users on TVs, with online microblogging services.

Profiling TV shows is important to both content providers and users [10]. On one hand, an interesting profile of a TV show will allow a content provider to demonstrate the TV show more effectively, e.g., more users can be attracted to watch the TV show,

indicating more subscriptions from the content provider. On the other hand, an illustrating profile of a TV show also gives a user an efficient way to know about the TV show before actually watching it, improving the whole show viewing experience.

Today, it has become a norm rather than an exception, for users to share and receive information and use online social networks to search multimedia contents [1]. In this personal and social context, even TV show profiles have to be highly personalized and social-aware. Traditional video profiling is generally based on video content only [3], e.g., the profile is composed by the video title, actor information, etc., and has the following fundamental drawbacks: (1) For a content provider, the video profile is generated without any consideration about the *social effect* the profile may make, e.g., how the video is going to attract users in the online social network and how the profile will be correlated with other videos. (2) For users, the profile is *static* and composed in an one-size-fit-all manner, i.e., the same video profile is provided to all users without any user-specific information.

To address these challenges, we propose a TV show profiling framework called MAP by exploring the new design space of using online social network information. Our contributions in this paper are summarized as follows: (1) We design a joint user and content information retrieval scheme. We use information about both actors and topics in a TV show to retrieve microblogs from a microblogging system. (2) We propose a social-aware profiling scheme, which profiles a TV show according to not only its content, but also the social relationship of its microblogging users and its propagation in the social network. (3) We present some interesting results and analysis, using our framework to profile real-world TV shows.

The rest of this paper is organized as follows. We present the information retrieval strategy and the microblogging-assisted profiling framework in Sec. 2. Based on this framework, we present several analytical results in Sec. 3, using data from real-world TV show and microblogging systems. Finally, we conclude the paper in Sec. 4.

2 Profiling TV Shows with Microblogs

In this section, we present our profiling framework. First, to effectively utilize the noisy microblog information, we retrieve the most informative microblogs for a TV show, by jointly considering actors and topics of the show; Second, to efficiently provide a profile both informative to a TV show content provider and users, we create a profile for TV shows, from the perspective of user, content, social relationship and propagation.

2.1 Background and Data Collection

We have collected data from a TV show service provider *BesTV* and an online microblogging service provider *Tencent Weibo* for our study. For simplicity, we refer to the TV show service as V, and the microblogging service as M. We have collected TV shows from the TV show company V. Our dataset contains 233,804,070 valid user viewing records of 69 TV shows during June 1 through December 30, 2011. We have also collected the meta information of these TV shows, including the title, category, and the actors in the TV shows. In Table 1, we list the representative TV shows used in our study. We will use the indices (v_1, v_2, \dots, v_{15}) to denote these TV shows in this paper.

Table 1. Representative TV shows studied.

ID	TV Show title	Category	Leading Actors	# microblogs	# microblog users
v1	<i>Aiqing Shuixingle</i>	Love/Idol	Stephy Qi/Tiffany Tang	571, 227	388, 109
v2	<i>Bubu Jingxin</i>	Love/Costume	Shishi Liu/Nicky Wu	2, 748, 223	1, 772, 973
v3	<i>Meiren Xinji</i>	Love/Costume	Ruby Lin/Mi Yang	169, 513	147, 812
v4	<i>Nanrenbang</i>	Idol/Modern	Di Yao/Lei Huang	1, 141, 362	844, 655
v5	<i>Qianshan Muxue</i>	Love/Idol	Yinger/Hawick Lau	271, 371	190, 288
v6	<i>Qingshi Huangfei</i>	Love/Costume	Ruby Lin	729, 010	544, 439
v7	<i>Xuebao</i>	War/Love	Zhang Wen	83, 835	67, 561
v8	<i>Wuxiekeji</i>	Love/Idol	Stephy Qi/Tiffany Tang	158, 534	116, 436
v9	<i>Aiqing Gongyu</i>	Comedy/Love	Loura/Jean Lee	1, 036, 144	775, 171
v10	<i>Aishang Zhameile</i>	Love/Idol	Cyndi Wong	78, 522	59, 080
v11	<i>Shachun</i>	Family/Tragedy	Helen Tao	56, 524	47, 582
v12	<i>Duanci</i>	War/Suspense	Lei Tong	14, 509	11, 410
v13	<i>Duanhounu</i>	War/Historical	Yunxiang Gao	5, 370	4, 260
v14	<i>Huangtuteng</i>	Love/Costume	Hans Zhang/Bing Bai	66, 881	48, 400
v15	<i>Luohun Shidai</i>	Love/Family	Zhang Wen/Di Yao	3, 544, 006	2, 789, 363

2.2 Informative Microblogs Retrieval

Microblogs are often intrinsically noisy, as users have been using such online microblogging services to exchange all kinds of information. Besides, among these microblogs posted by users, only a little fraction of them are about TV shows, e.g., in popular microblogging systems including Twitter, only a very small fraction of microblogs are about TV shows¹. It is a must to find the most informative contents when we use microblogs to profile TV shows. We propose to jointly use both actors and topics of a TV show to retrieve the most informative microblogs.

Microblogs from Actors. In our study, we focus on the professional TV shows, which are produced by large companies, instead of user-generated videos [7], as such professional videos (e.g., a movie or TV show) are the mainstream type of videos on a video service. For such a TV show v , we are able to find an actor list \mathbf{A}_v , in which actors of the show are included. According to our measurement results, there is over 1/3 of the leading actors having a microblogging account on average. We collect microblogs posted by accounts in \mathbf{A}_v , if they have social accounts. For example, a TV show v_6 in our study named “*Qingshi Huangfei*” has an actor named “Ruby Lin”, who has a social account ID on the microblogging system M called “linxinru”, and we have collected microblogs posted by this ID for profiling this TV show.

Microblogs from Topics. For a TV show v , we are also able to find a set of topics \mathbf{O}_v , including the title of the TV show, category, abstract, etc. In our study, such topics are used to search for microblogs, in a keyword-matching manner, i.e., if the content

¹ <http://en.wikipedia.org/wiki/Twitter>

of a microblog contains any keyword in \mathbf{O}_v , this microblog will be used for profiling TV show v . For example, the same TV show v_6 “*Qingshi Huangfer*” has the leading roles “Mafuya”, “Mengqiyou”, “Liuliancheng”, which are used as the topics to retrieve microblogs.

Using the above microblogs as “seed” microblogs, we further collected microblogs that are repost and comments to them. In summary, we have collected overall 23,963,680 valid microblogs posted by 11,685,768 users in M. In Table 1, we show the number of microblog users and microblogs for each of the representative TV shows. Each microblog record contains the follows ID, name, IP address of the publisher, time stamp when the microblog was posted, IDs of the root microblog if it is a reposted microblog, and contents of the microblog. Interested readers are referred to our previous studies on collecting the microblog traces [11,12].

2.3 MAP: Microblogging-Assisted Profiling of TV Shows

Based on microblogs, user profiles, and social connections collected using the retrieval scheme above, we profile TV shows from the following aspects. (1) *User aspect*: We present the demographics of microblogging users of a TV show, and their preferences; (2) *Content aspect*: We present the sentiment analyses of TV shows by the microblog content posted to the shows, and the “social network” of TV shows; (3) *Social relationship aspect*: We present the social topology of microblogs views of a TV show, as well as the social network and social influence of actors of TV shows, who have online microblogging accounts; (4) We present the propagation of microblogs users across different TV shows, i.e., how users’ attention is shifting from a TV show to another.

Next, we present a case study using MAP, including the insights from the observations.

3 Insights Learnt from MAP

In this section, we present several results based on our profiling framework, from the aspects of user, content, social relationship and social propagation.

3.1 Demographics and Preference of Viewers

Demographics of Microblogging Viewers. Using the microblogging user information, we are able to demonstrate TV shows from a user aspect. Fig. 1(a) shows the age distribution of normal users. It forms a normal-like distribution, and we observe that the average age is 20, indicating the major of the users are quite young. Fig. 1(b) shows the top 15 province-level regions where the microblog users are located. Each bar in the figure is a top-10 participation index, defined as

$$PI_i = \frac{UN_i - UN_{10}}{UN_{10}} \quad (1)$$

where UN_i is the numbers of users in region i . From this demographics profiling, we are able to show the regional distribution of users. For example, we observe that for this particular TV show provider, Guangdong Province ranks the first, i.e., it has the most microblog users. In summary, we observe that the young users in top-tier cities in China are the major viewers of these TV shows.

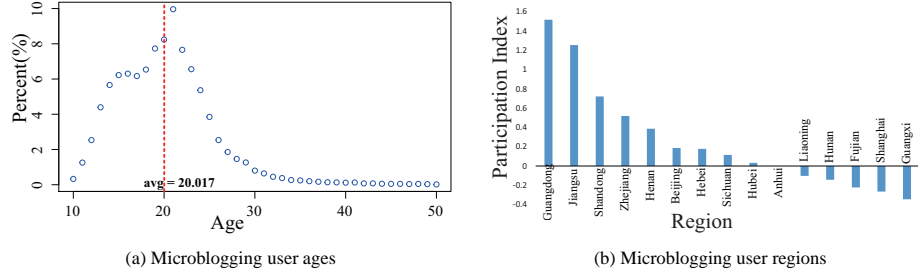


Fig. 1. Demographic of users.

User Preference Clustering. A TV show provider gives 3 labels to each TV show to categorize its content. We use such content labels to infer users' preferences of video contents. According to the microblogs posted by a user, we are able to assign each user with a weight list of content labels. Using the weighted label lists of users, we are able to calculate a "distance" between any two users. Then, we cluster these users using a K-means clustering algorithm [5] using based on the label distance.

In Fig. 2(a), we plot the social connections between users that clustered into one group. Each circle represents a user, and a connection between two users indicates that one of them is following the other. We observe that in this figure users are closely connected to each other with a large clustering coefficient [9], when they are clustered into the same group by K-means according to their preference.

Using the same approach to cluster users, we further find that users' preference can be indicated from who they are socially connected in a microblogging system. In Fig. 2(b), we have plotted several VIP microblog users, i.e., *Tencent Video*, *Hunan TV*, and *Tencent Entertainment*. We observe that when users are clustered into the same group, they only follow the VIP users in that group, but have no connection with other users in the same group. This observation indicates that a TV show provider is able to identify users that may be interested in a TV show, by looking at the followers the VIP accounts.

3.2 Sentiment and Social Network of TV Shows

Online microblogging services also allow us to explore the sentiment and social network of TV shows.

Sentiment Analysis. Microblogs allow us to use the content of microblogs for sentiment analysis, i.e., we label each microblog "positive", "negative", or "non-sentiment",

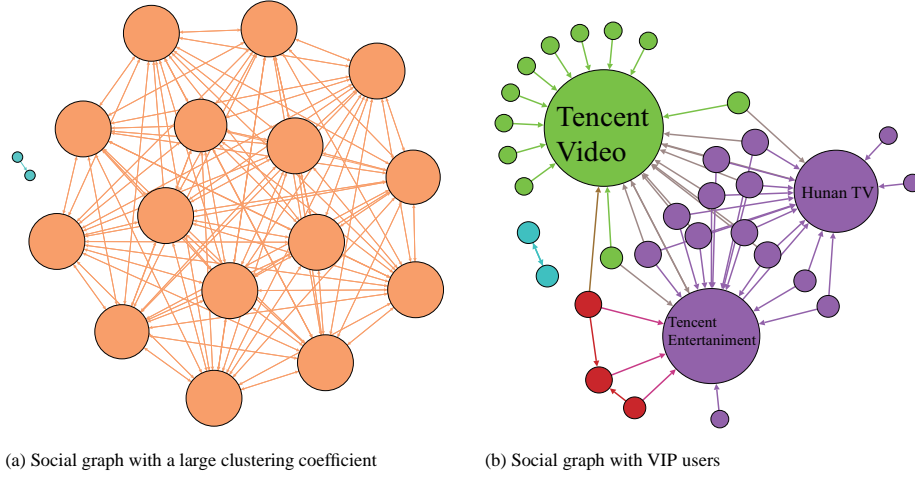


Fig. 2. Social relationship among users sharing similar preference.

according to its content. After analyzing all the posts, we get the following observations: (1) Initial microblog posts on TV shows contain more emotional contents than reposts and comments, as illustrated in Fig. 3(a); (2) There are much more positive microblogs on TV shows than negative ones, as illustrated in Fig. 3(b).

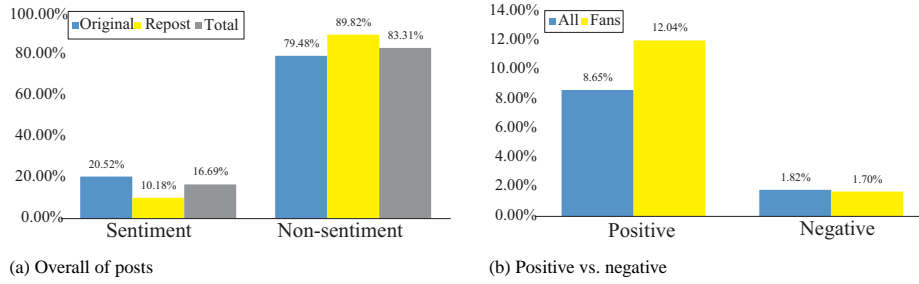


Fig. 3. Sentiment distribution of microblogs on TV shows.

Next, we study the sentiment distribution of microblogs in different TV shows. We find that most of them have higher positive proportions except for two TV shows. As shown in Fig. 4, each bar is the fraction of positive microblogs versus the TV show rank, and the curve is the view number of each TV show versus the show rank. We observe that a TV show that has too many positive microblogs or too many negative microblogs tend to have a small number of views. The reason may be that highly-viewed shows will attract more diverse discussions on the microblogging system.

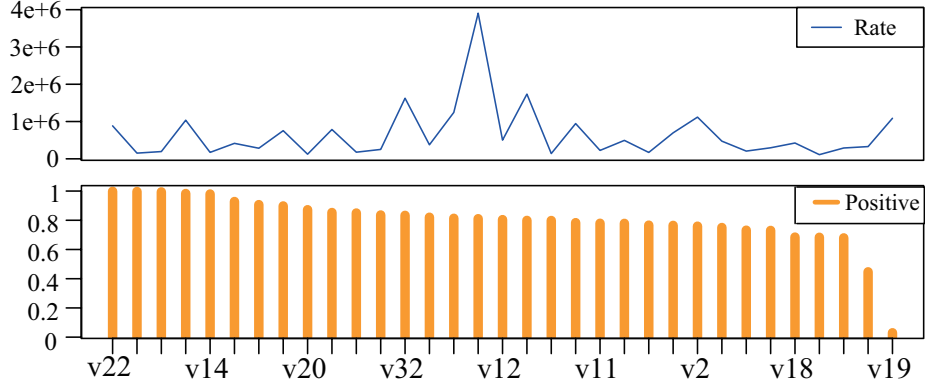


Fig. 4. Sentiment distribution of different TV shows, along with their TV rating.

“Social Network” Between TV Shows. Users can post microblogs which relate to more than one TV shows, therefore, TV shows may be connected via their common microblog users. Using this idea, we create a social network of TV shows, i.e., two shows are connected if they share a microblog user. Fig. 5 shows the statistics of nodes (TV shows). Fig. 5(a) illustrates the degree distribution of TV shows. We observe that about 90% of the shows have a degree higher than 40, and the mean is 45.8, indicating that the graph is almost fully-connected. Fig. 5(b) illustrates the CDF of local clustering coefficients of the shows. We note that TV show graphs with higher level clustering coefficient indicate more common users.

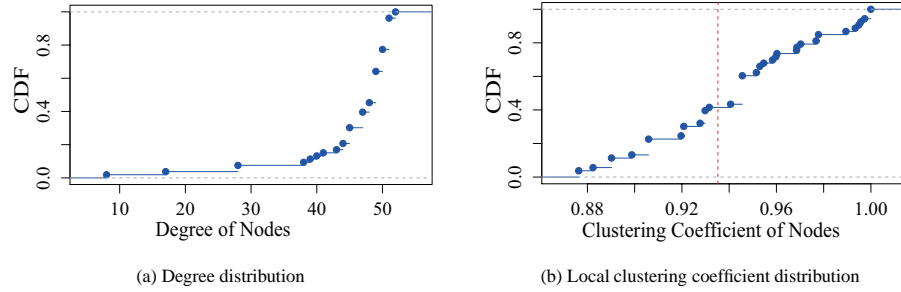


Fig. 5. Statistics of Nodes.

Next, we study the community structure of the social network of TV shows, as illustrated in Fig. 6. Since the average clustering coefficient of the network is 0.94, which is 2x of the same scale of a random network. The average diameter of the graph is 1.12, which is at the same scale of a random network. Also, we study its modularity using a heuristic approach [2]. We have modularity value of 0.05 (Using the community structure detecting algorithm proposed by Du et al. [4]) for the same social network of the

TV shows. The very small value indicates that the social network of TV shows does not have a clear community structure [8].

Furthermore, we explore the degree distribution of the network. Based on the topology, we have that the degree distribution of nodes in the TV show social network follows a power-law distribution $f(x) = -52x^{-0.5} + 58$ ($R^2 = 0.98$), indicating that the TV show social network can be regarded as a scale-free network².

3.3 Social Networks of Viewers and Actors

An online microblogging service also allows us to explore the social relationship between the microblog users who have posted microblogs on the TV shows.

Social Topology Among Viewers. We first study the following relationship among microblog users. As shown in Fig. 7, most of the users (about 76%) are isolated (with no social connection to other users), and a small fraction (about 24%) of them have social connections.

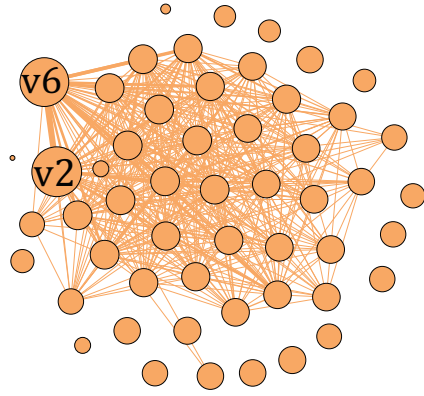


Fig. 6. Communities in the social network of TV shows.

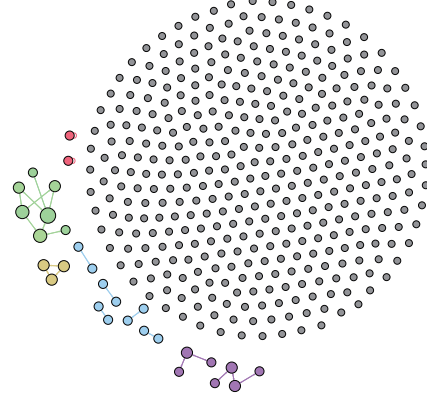


Fig. 7. Social graph among users of TV show v1.

Social Networks among Actors. Today, online microblogging services are not only used by ordinary users, but also popularly used by celebrities and actors. We next study the social networks of actors of TV shows. Fig. 8(a) is a graph created based on microblogs posted by actors: there is an edge between two actors if one of them has posted a microblog on a TV show which involves the other. In this figure, we observe that some actors tend to help others to broadcast their shows, even they are not acting in those shows. In Fig. 8(b), we have plotted the following relationship between actors on

² http://en.wikipedia.org/wiki/Scale_free_network

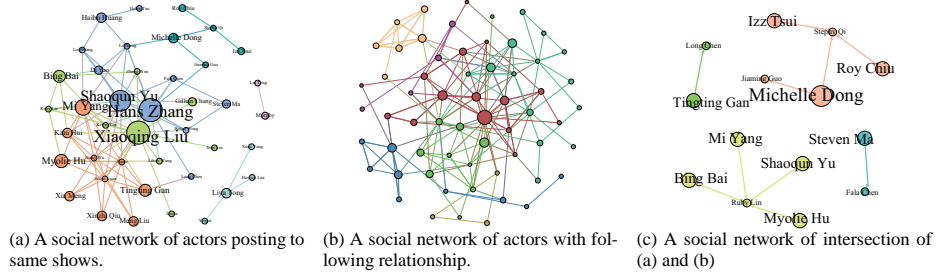
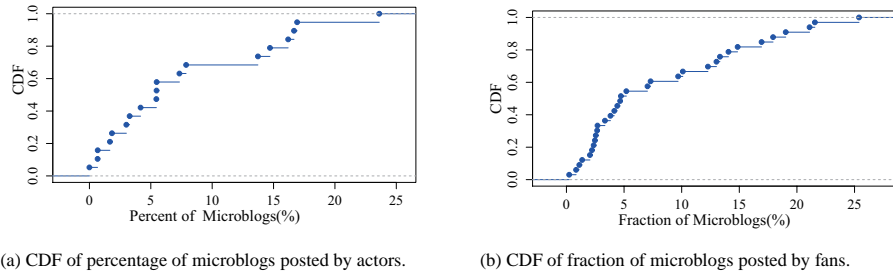


Fig. 8. Social networks among actors.

the microblogging system. We observe that actors also share similar social topologies as ordinary users [1]. In Fig. 8(c), we plot an intersection of (a) and (b), i.e., only edges that appear in both graph (a) and (b) are plotted. We observe very few edges remaining in this figure, indicating that it is very common that actors post microblogs on shows of other actors they are not following or friending on a microblogging system.

Social Influence of Actors. Next, we study the social influence of actors to the shows they star in.

Fig. 9(a) shows the CDF of the fraction of microblogs by actors over all microblogs about a TV show. We have the following observations: (1) Actors post a large fraction of microblogs in a TV show; (2) Though more than half actors post more than 5% TV-related microblogs, there is almost no actor posting more than 20%. Next, we study the microblogs posted by fans of the actors of a show. Fig. 9(b) plots the CDF of fractions of microblogs posted by fans of actors in a TV show. We observe that the fans of actors also input a large portion of microblogs for a TV show.



(a) CDF of percentage of microblogs posted by actors. (b) CDF of fraction of microblogs posted by fans.

Fig. 9. Social influence from actors of TV shows.

Finally, we study the heterogeneous influence of different actors. We calculate actors' average and variance influence on the shows they act in. The results are presented in Fig. 10, with their fans number. (1) We observe that for most of the actors, they tend to

have a higher average influence if they have a larger number of fans. (2) The influence variance is generally small.

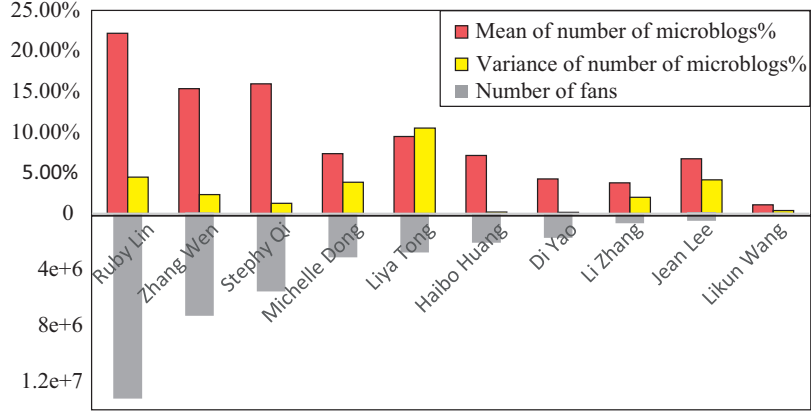


Fig. 10. Actors' social influence and number of fans.

3.4 Viewer Propagation across TV Shows

Viewers of TV shows generally enjoy different shows at different periods. Based on our profiling scheme, we are able to study the propagation of users across different shows.

Propagation of Viewers Over Time. We study how users join the same show played at different times. In this experiment, shows are played in two different periods, denoted as the first round and second round. In Fig. 11, each bar plots the number of users in the first round, second round, and both rounds. We observe that though the number of users in both rounds accounts for a small proportion. This observation indicates that new users tend to join a TV show if the show is played again.

Propagation of Viewers Across Shows. Next, we study user propagation between TV shows. In Fig. 12, each node represents a TV show, and its diameter indicates the number of in-degrees. There is a directed edge between two nodes if users “propagate” from one TV show to another, i.e., microblogging users start to post microblogs to another from one TV show. The thickness of an edge indicates the number of users being propagating. In Fig. 12(a), we study a propagation case when a TV show v2 has a plummeting number of microblogging users. We observe that much more social-network users are propagating from v2 to other shows. Besides, we observe that among these propagation users, a major fraction of them are moving to TV show v1. The reason

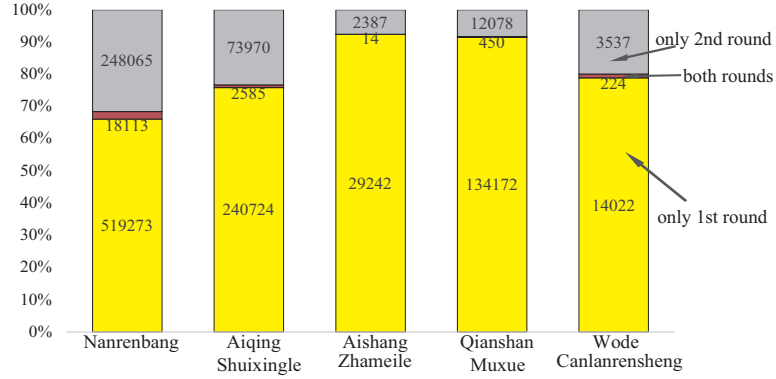


Fig. 11. Propagation of viewers over time.

is that the content provider V began to promote v1 two days before the propagation, indicating that users can be highly influenced by new TV shows published.

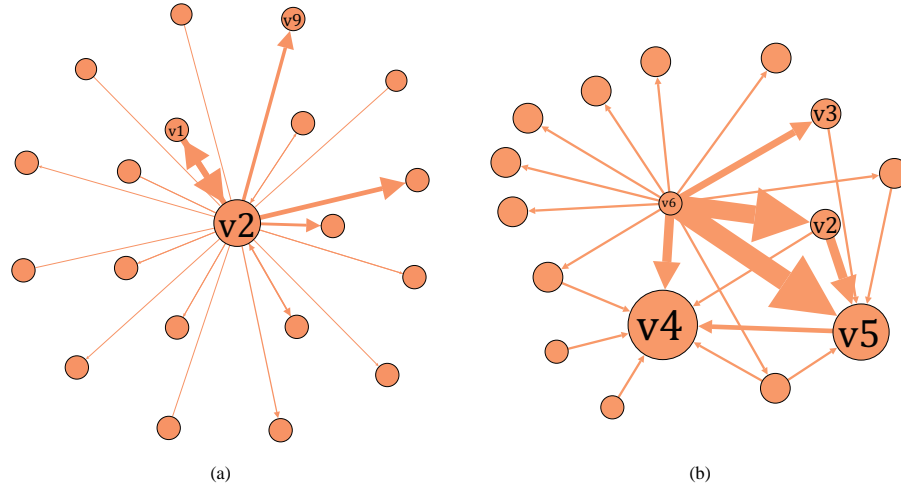


Fig. 12. Propagation of microblogging users between TV shows.

In Fig. 12(b), we study another type of propagation. This figure illustrates a case when v6 is about to finish (delivering its last episode) and v4 and v5 are the newly released shows scheduled on the next day. We observe that a dominate fraction of users are propagating from v6 to v5. The reason is that v6 and v5 are in the same TV category. On the other hand, v4 is of a different TV category, and we observe that much fewer users are propagating from v6 to v4. This observation indicates that users tend to propagate between TV shows with similar content.

4 Concluding Remarks

In this paper, we study the profiling of TV shows using online microblogging services, in a sense to provide users with a socialized profile of TV shows. It is challenging to profile TV shows with microblogs since TV shows are generally offline, and an effective information retrieval scheme is in demand to find the most useful microblogs for the profiling; meanwhile, information on a microblogging system is noisy for the profiling and an efficient profiling scheme is in demand. To address these two challenges, we propose a joint actor and topic information retrieval scheme for searching microblogs to profile a particular TV show. Based on that, we propose *MAP*, a microblog-assisted profiling framework, in which a TV show is profiled from the aspects of user, content, social relationship and social propagation. Using *MAP*, we profile real-world TV shows and present several analysis results, which are not only interesting to the ordinary viewers, but also important to content providers for better operation.

Acknowledgments. This research is supported in part by NSFC under Grant No. 61272231, 61472204, and 61402247, SZSTI under Grant No. JCYJ20140417115840259, and Beijing Key Laboratory of Networked Multimedia. We thank Tsinghua-Tencent Joint Lab and BesTV for providing the traces used our study.

References

1. Benevenuto, Fabrício and Rodrigues, Tiago: Characterizing user behavior in online social networks. Proceedings of the 9th ACM SIGCOMM. ACM, 49-62.(2009)
2. Blondel, Vincent D and Lambiotte, Renaud: Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. 2008(10):P10008.(2008)
3. Dimitrova, Nevenka and Zhang, Hongjiang: Applications of video-content analysis and retrieval. IEEE MultiMedia. 9(3):4255.(2002)
4. Du, Haifeng and Feldman, Marcus W and Li, Shuzhuo and Jin, Xiaoyi: An algorithm for detecting community structure of social networks based on prior knowledge and modularity. Complexity, 12(3):5360.(2007)
5. Hartigan, John A and Wong, Manchek A: Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc., 28(1):100108.(1979)
6. Holotescu, Carmen and Grosseck, Gabriela: M3-learning-Exploring mobile multimedia microblogging learning. World Journal on Educational Technology, 3(3):168176.(2011)
7. Milliken, Mary C and Gibson, Kerri: User-generated video and the online public sphere: Will YouTube facilitate digital freedom of expression in Atlantic Canada?. American Communication Journal, 10(3):114.(2008)
8. Newman, Mark EJ.: The structure and function of complex networks. SIAM review, 45(2):167256.(2003)
9. Soffer, Sara Nadiv and Vázquez, Alexei: Network clustering coefficient without degree-correlation biases. Physical Review E, 71(5):057101 14.(2005)
10. Wakamiya, Shoko and Ryong, LEE: Twitter-based TV Audience Behavior Estimation for Better TV Ratings. DEIM Forum. (2011)
11. Zhi Wang and Lifeng Sun and Xiangwen Chen and Wenwu Zhu and Jiangchuan Liu and Minghua Chen and Shiqiang Yang: Propagation-based Social-aware Replication for Social Video Contents. ACM International Conference on Multimedia (Multimedia). (2012)
12. Zhi Wang and Lifeng Sun and Wenwu Zhu and Shiqiang Yang and Hongzhi Li and Dapeng Wu: Joint Social and Content Recommendation for User Generated Videos in Online Social Network. IEEE Transactions on Multimedia, 15(3):698709. (2013)
13. Xu, Tianyin and Chen, Yang: Cuckoo: towards decentralized, socio-aware online microblogging services and data measurements. Proceedings of the 2nd ACM HotPlanet. ACM.(2010)